Draft NIH Genomic Data Sharing Policy Public Consultation Webinar

Wednesday, November 6, 2013

Draft NIH Genomic Data Sharing Policy: Introduction

Amy P. Patterson, M.D.
Associate Director for Science Policy, NIH

Overview of Webinar

- Introduction
- NIH and Data Sharing: A Quick Recap
- Overview of the Draft GDS Policy
- Question and Answer Session
- Wrap Up

Data Sharing Supports NIH's Mission and Priorities

- Data sharing supports the NIH mission by maximizing knowledge by:
  - Enabling data generated from one study to be used to explore a wide range of additional research questions
  - Increasing statistical power and scientific value by enabling data from multiple studies to be combined
  - Facilitating validation of research results
  - Facilitating innovation of methods and tools for research
  - Ensuring the ethical conduct of research

Impetus for Extension of Data Sharing Policy

- Generation of larger volume of genomic data
  - NIH able to fund research that generates larger volume of GWAS and other types of genomic data due to advances in DNA sequencing and other high throughput technologies and a steep drop in sequencing costs
- Calls for expanded data sharing from public and private sectors
  - February 2013 White House initiative to increase access to the results of federally funded scientific research
  - NIH Big Data to Knowledge Initiative (BD2K)
  - Proposed Common Rule revisions (ANPRM, July 2011) supports broad consent to maximize utility of biospecimens and data
- Respect participant interests and wishes

NIH and Data Sharing: A Quick Recap

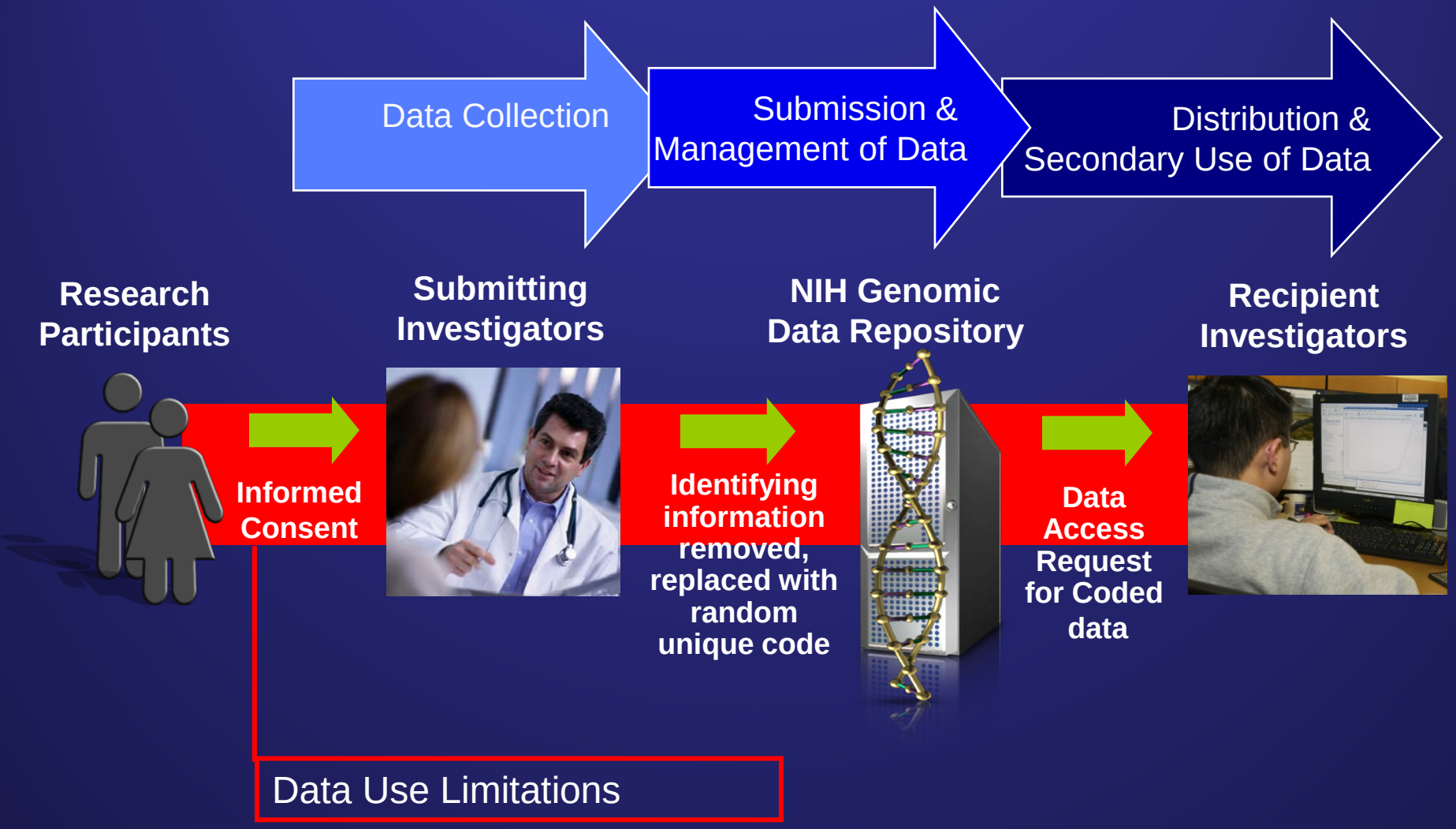
Laura Lyman Rodriguez, Ph.D.
Director, Division of Policy, Communications, and Education
National Human Genome Research Institute, NIH

A Culture of Sharing

NIH has long history of promoting data sharing:

- 1999 Research Tools guidelines
- 2001 Grants Policy Sharing Guidance
- 2003 Data Sharing Policy
- 2004 Model Organism Policy
- 2007 GWAS Policy

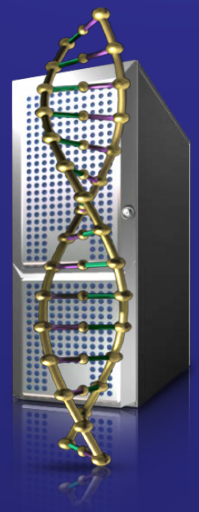
Genomic Data Management Overview



GWAS Data Access is Two-Tiered

Genotype & Phenotype Data

dbGaP—NIH Genomic Data Repository



Public Access

Study Protocol
Descriptive Information

Controlled Access

Coded Genotypes
Phenotypes
Pre-computes

- Review data use limitations

Data Access Committee

All potential users

- Co-signed by institution
- Agree to terms of use
- PI agree to Code of Conduct

Requested Research Use

dbGaP Data Access and Use 2007- April 2013

Total Studies Available in dbGaP = 407

Data Access Requests (DARs) = 16,927
(Approved DARs = 11,638)

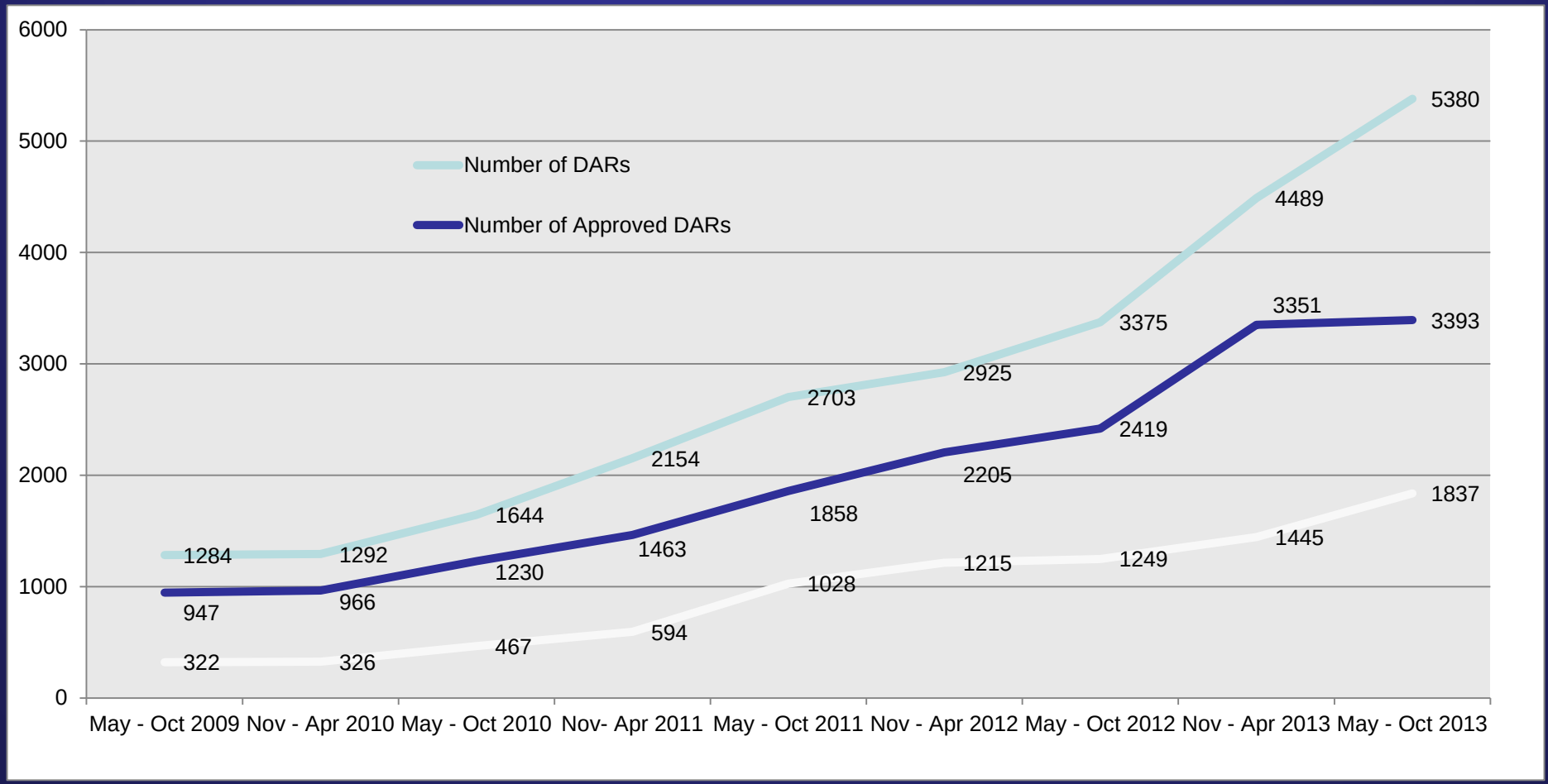
Most Common Research Uses

- **Statistical methods research**
- **Software development**
- **Developing medical therapies**
- **Basic scientific investigations**

Secondary user publications = ~450/yr

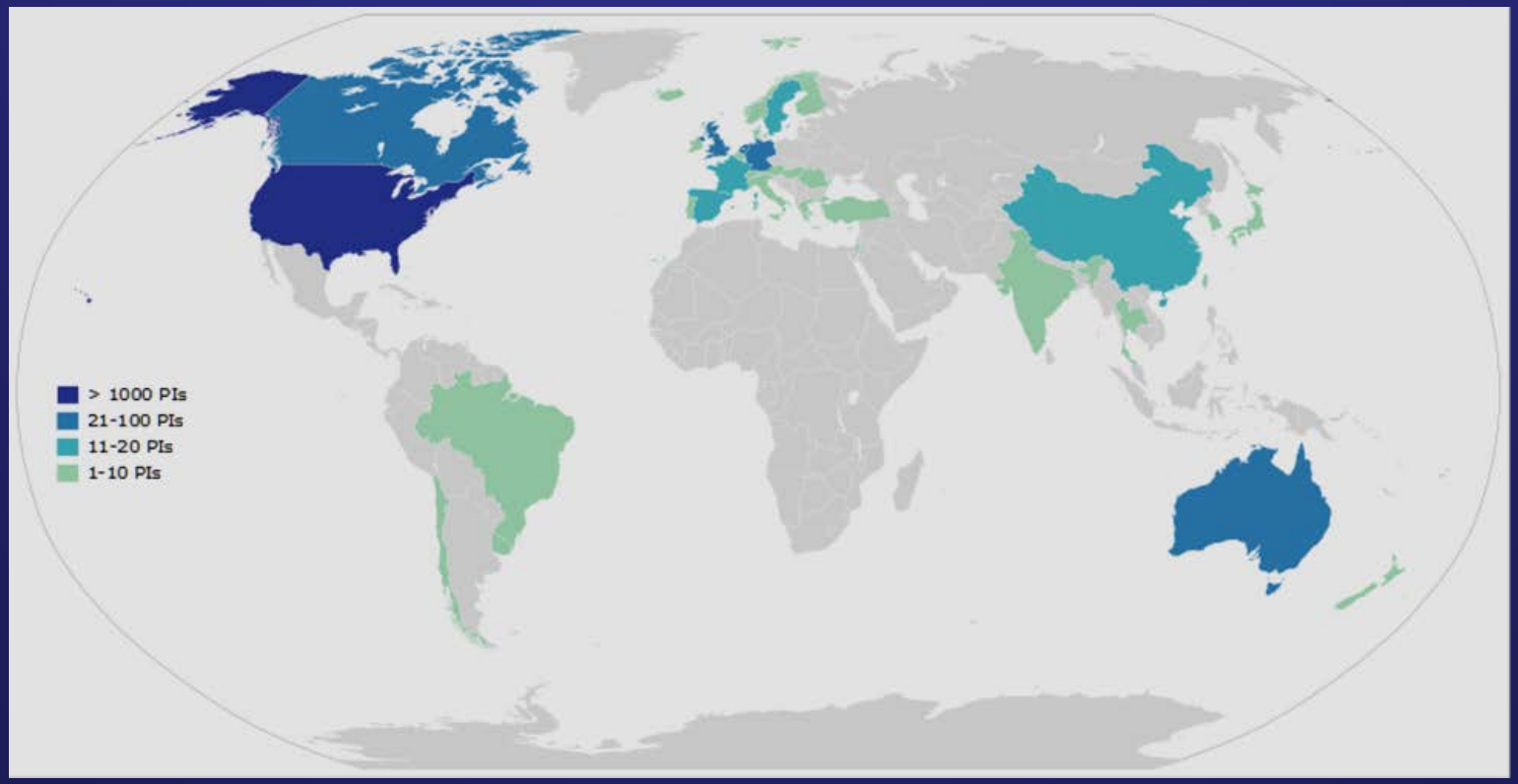
(Based on last 1.5 years)

Trends in dbGaP Data Use



Range of dbGaP Data Users

- Approved users from 36 countries
- 2,118 investigators approved from 7/2007 – 11/2013
- > 500 organizations across the research community



Need for a Broader Policy

- Whole genome analysis represents a pathway through which to advance understanding of common diseases (e.g., diabetes, cancer, heart disease)
- The data generated are far richer than what a single investigator or a collaborative team can fully explore
  - Many different questions may be asked
  - Cross-study analyses are possible, which increases the capacity to address complex questions
- Extends participant protections beyond GWAS to other types of genomic data
- An overarching policy framework that promotes consistent and robust data sharing can best serve the public

Overview of the NIH Draft Genomic Data Sharing Policy

Dina Paltoo, Ph.D., M.P.H.
Director, Genomics, Health, and Society Program
Office of Science Policy, NIH

Draft GDS Policy – Main Components

- Scope and applicability
- Effective date
- Responsibilities for investigators submitting genomic data, including:
  - Data sharing plans
  - Data submission expectations and timeline
  - Data repositories
  - Informed consent
  - Institutional certifications
  - Exceptions to expectations
- Responsibilities of investigators accessing and using genomic data, including:
  - Requests for controlled-access data
  - Acknowledgment for use data
- Intellectual Property

Draft GDS Policy – Scope and Applicability

- All NIH-funded research involving non-human organisms or human specimens that produces genomic, metagenomic, epigenomic, or transcriptomic data from large-output sequencing instruments or genotyping platforms, such as:
  - Sequence data from tens of isolates from infectious organisms
  - Sequence data from more than one gene or gene-sized region in more than 100 participants
  - Data from more than 10,000 genes or regions from one participant (e.g., whole genome sequencing)
  - Data from more than 100,000 variant sites in more than 100 participants
- Applies to all funding mechanisms (grants, contracts, or NIH intramural support); no minimum threshold for cost
- NIH will periodically review the scope of the GDS Policy and make changes as necessary

Effective Date of the Final GDS Policy

- Effective in 2014, 60 days after publication of final GDS Policy

- Implemented for research proposals submitted in 2015 for FY2016 funding

Draft GDS Policy – Responsibilities of Investigators Submitting Data

- PIs seeking funding should contact appropriate NIH Project Officials to discuss expectations and timelines for sharing
- Plans for conforming with the GDS Policy should be included in the data sharing section of funding applications
  - Plans should include resources necessary to support sharing
- NIH intramural PIs should contact their IC leadership or OIR for guidance

Key Distinctions between GWAS and GDS

	GWAS Policy	GDS Policy
Scope	Applies to human GWAS data	Applies to all genomic data types, human and non-human
Consent Standard -- Existing* Collections **Before the effective date of the Policy*	If research consent, IRB reviews for consistency. If no research consent exists, data may still be submitted to NIH databases.	Same
Consent Standard -- Future* Collections **After the effective date of the Policy*	Policy is silent but the preamble states "the NIH expects specific discussion within the informed consent process and documentation that participants' genotype and phenotype data will be shared for research purposes through the NIH GWAS data repository."	Samples or cell lines should be consented for research use and broad data sharing. Exceptions can be requested. Consent should address whether data should be shared in open or controlled access.
Data Submission	Data submitted as soon as quality control procedures are completed	Timelines vary by data type, but generally as soon quality control procedures are complete
Data Release	Immediate data release. 12 month publication embargo	6 month deferral of data release. No publication embargo

Draft GDS Policy – Non-human Genomic Data

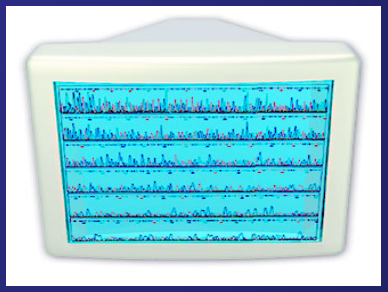
- Encourages consistent data sharing practices
- Expectation for sharing is consistent with current practice and recent Federal policy initiatives
  - e.g., NIH Model Organism Policy, White House initiative
- Current resources and databases will remain the standard mechanism for sharing
- Flexibility for ICs to adjust expectations for research programs for different data types or types of projects
  - *e.g.*, microbial data pre-publication, model organism data no later than publication

Management of Controlled-Access Human Genomic Data

Data Submission

Informed Consent

De-identified Data

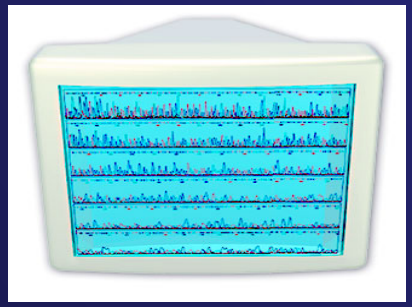


Research Participants

Submitting Investigator and Institution

NIH Genomic Data Repository

Data Access



Data Access Request

Access Granted for One Year

NIH Genomic Data Repository

NIH Data Access Committee

Recipient Investigator

Data Use Limitations

Draft GDS Policy – Informed Consent Expectations

- Expectation of explicit consent for broad sharing for research use of data from specimens or cell lines collected or generated after the Policy's effective date
  - Informed consent process and documents should address whether data are to be shared in open access or controlled access
  - Consent expected for clinical specimens and cell lines collected or generated after effective date, even if de-identified, unless there are compelling scientific reasons

- Allows for continued use of data from clinical specimens or cell lines collected or generated before the Policy's effective date
  - For specimens collected before the effective date, assessment by an IRB should assure that data submission is not inconsistent with the informed consent

Draft GDS Policy – Responsibilities of Institutions Submitting Data

- A Signing Official from the institution submitting data provides an Institutional Certification assuring:
  - The data submission is consistent with laws, regulations, and institutional policies;
  - The appropriate research uses of the data and any uses that are excluded in the informed consent documents are delineated;
  - The identities of research participants will not be disclosed to NIH-designated data repositories; and
  - An IRB has reviewed the investigator's proposal to submit data

Draft GDS Policy – Responsibilities of IRBs Reviewing Submissions

- Prior to submission, the IRB reviews the proposal to assure:
  - The protocol for collection of specimens or data was consistent with Federal regulations for human subjects research;
  - Submission and sharing of data for research are consistent with the informed consent;
  - Risks to individuals and their families, and groups or populations, associated with data submitted to NIH-designated repositories were considered;
  - PI's plan for de-identifying datasets is consistent with the GDS Policy's standards
- For studies using specimens or data collected before the effective date, NIH expects an assessment by an IRB prior to submission

Draft GDS Policy – Expectations for Data Submission and Data Release

Level	General Description of Data Processing	Example Data Types	Data Submission Expectation	Data Release Timeline
0	Raw data generated directly from the instrument platform	Instrument image data	Not expected	NA
1	Initial sequence reads, the most fundamental form of the data after the basic translation of raw input	DNA sequencing reads, ChIP-Seq reads, RNA-Seq reads, SNP arrays, arrayCGH	Not expected for human data if reads are included in Level 2 aligned sequence file (e.g., BAM) Non-human de novo sequence data	NA Up to 6 months for non-human data
2	Data after an initial round of analysis or computation to clean the data and assess basic quality measures	DNA sequence alignments to a reference sequence or de novo assembly, RNA expression profiling	Project specific, generally within 3 months after data generation	Up to 6 months after data submission or at the time of acceptance of the first publication, whichever occurs first
3	Analysis to identify genetic variants, gene expression patterns, or other features of the dataset	SNP or structural variant calls, expression peaks, epigenomic features	Project specific, generally within 3 months after data generation	Up to 6 months after data submission or at the time of acceptance of the first publication, whichever occurs first
4	Final analysis that relates the genomic data to phenotype or other biological states	Genotype-phenotype relationships, relationships of RNA expression or epigenomic patterns to biological state	Data submitted as analyses are completed	Data released with publication

Draft GDS Policy – Responsibilities of PIs Accessing/Using Data

- In order to download or use controlled-access data, PIs must request access
- PIs approved to download agree to a Data Use Certification cosigned by their Institutional Signing Official that includes terms and conditions such as:
  - Using the data only for the approved research
  - Protecting data confidentiality
  - Following applicable laws, regulations, and policies for data use
  - Not attempting to re-identify individual participants
  - Not sharing the data with individuals not listed in the data access request
  - Agreeing to report violations of the GDS Policy to the appropriate NIH data access committee immediately
  - Providing annual updates to NIH on research

Draft GDS Policy – Responsibilities of PIs Accessing/Using Data (cont'd)

- PIs also agree to abide by the NIH User Code of Conduct
- In any publication, presentation, or other public reporting of results of research using data accessed through NIH-designated data repositories, PIs agree to acknowledge:
  - The original PI that submitted the data;
  - The funding organization(s) that supported the original research;
  - The dataset and its accession number (e.g., phs000###); and
  - The NIH data repository through which the data were accessed

Intellectual Property

- Supreme Court has found that naturally occurring DNA sequences are not possible in the U.S.
- Draft GDS Policy considers basic sequence data and related information (genotypes, haplotypes, *p*-values, allele frequencies) in NIH-designated repositories to be “pre-competitive”
- These data and any conclusions derived from them should remain freely available without licensing requirements
- Discourages the use of patents to block access to genotype-phenotype data

Draft GDS Policy – Public Consultation Process

- Notice of the draft Policy was published in the Federal Register (Document number 2013-22941) and the NIH Guide (NOT-OD-13-119)
  - 60-Day public comment period ending November 20
  - Accepting comments on any aspect of the draft Policy
  - Web interface for gathering public comments: Public webinar with a question and answer session

- Comments will be posted on the GDS website at the end of the comment period

QUESTION AND ANSWER SESSION

Importance of Public Comments

- The public comment periods provides an opportunity for NIH to:
  - Learn perspectives of a broad range of stakeholders (e.g., investigators, research participants, IRBs, ethicists)
  - Consider concerns about specific aspects of the GDS Policy
  - Use public feedback to shape the final Policy

THANK YOU!

Please Submit Public Comments by Wednesday, November 20, 2013

Go to http://gds.nih.gov for more information